A NEW INTEGRATED METHOD FOR SHAPE BASED IMAGE RETRIEVAL

Xiaojun Qi and Hongxing Zheng
Computer Science Department, Utah State University, Logan, UT, 84322-4205
xqi@cc.usu.edu and hzheng@cc.usu.edu

ABSTRACT

This paper proposes a novel and efficient shape retrieval scheme, which is robust to RST (Rotation, Scaling, and Translation). The proposed approach integrates global and local shape descriptors for accurate retrieval, where the global descriptors are obtained from the Fourier transformation, and the local descriptors are obtained from a one-level wavelet transformation. The global and local similarity scores for each query and candidate image are individually computed using different measures. A Gaussian-based fuzzy method then combines the global and local similarity scores into one similarity membership, which measures the overall similarity between the query and candidate images. Experiments are performed on four different databases and the retrieval results demonstrate that the proposed approach not only accurately and efficiently retrieves shapes but also outperforms some peer systems in the literature.

KEY WORDS

Fourier descriptor; wavelet transform, and log-polar mapping.

1. Introduction and Related Work

Shape represents the contour of the objects in an image and is one of the most important human perceptual features. Shape-based retrieval has been prevalently used in pattern recognition and CBIR (Content-Based Image Retrieval) of digital images. Such retrieval normally requires a shape representation invariant to RST.

Many approaches have been proposed to perform CBIR using shape. Zhang and Lu [1] use FDs (Fourier Descriptors) of different shape signatures to study the shape retrieval results. Their experimental results demonstrate that the FD-based methods capture the general shape of the image without using any local feature. A context-based shape descriptor is proposed [2] to represent each point of the shape. The descriptor of each point captures the distributions of its related points. The similarity between two shapes is defined as the sum of the matching errors of the corresponding descriptors, together with the alignment transformation magnitudes. WTMM (Wavelet Transform Modulus Maxima) is used [3] to locate the high curvature points at different wavelet decomposition levels. The similarity score between two images is calculated by comparing each corresponding high curvature point after aligning the highest wavelet modulus magnitude points. This alignment provides the capability of the RST invariance. However, the alignment error may seriously affect the retrieval accuracy. Chen and Wang [4] propose a fuzzy method for region-based image retrieval, in which an image is represented by a set of segmented regions, and each region is characterized by a set of fuzzy features. The similarity between two images is defined as the overall similarity between two families of fuzzy features. In general, either the complexity of the above algorithms increases exponentially as the number of candidate images increases or the algorithms have long feature vectors at each dyadic scale for approximate comparison. As a result, these techniques are not suitable for large image databases with thousands of images.

In this paper, we propose a FD-and-WTMM-based fuzzy method for an efficient RST-invariant non-occluded shape retrieval. Our digital image data exclusively contains one clean non-occluded object boundary per image. The proposed method also utilizes the invariance properties of the LPM (Log-Polar-Mapping) [5] to ensure the scaling and rotation invariance. The integration of LPM/ILPM (Inverse Log-Polar-Mapping), FD, and WTMM not only computes the global and local descriptors, but also ensures the RST invariance. A Gaussian-based fuzzy method is further used to combine the global and local similarity scores into one membership, which measures the overall similarity between the query and candidate images by compensating the possible errors from either similarity measure.

The rest of the paper is organized as follows: Section 2 briefly introduces several concepts used in this paper. Section 3 proposes the novel integration approach to shape retrieval. Section 4 illustrates the experimental results. Section 5 draws conclusions.

2. Several Concepts

2.1 Log-Polar Mapping

In LPM, pixels are indexed by the log-polar coordinates (*R, W*)'s, where *R* is the ring number and *W* is the wedge number [5]. They are mapped from point (x, y) in the Cartesian coordinates into the log-polar coordinates by (1):

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2}\,; \theta = \tan^{-1}\frac{y - y_c}{x - x_c}$$

$$R = \frac{(N_r - 1)\ln\left(\frac{r}{r_{\min}}\right)}{\ln\left(\frac{r_{\max}}{r_{\min}}\right)}; W = \frac{Nw\theta}{2\pi} \quad (1)$$

where:

- (x_c, y_c) is the Cartesian coordinate used as the center of the lop-polar sampling (i.e., the origin of the log-polar coordinates);
- N_r is the number of rings (i.e., the sampling rate at the radial direction);
- N_w is the number of wedges (i.e., the sampling rate at the angular direction);
- $r_{\min}$ and $r_{\max}$ are the radii of the smallest and largest ring of samples.

Its main properties include scaling invariance and the conversion of a rotation in the Cartesian coordinates into a cyclic shift in the log-polar coordinates. The scale of an image in the log-polar coordinates is not changed, if the sampling rates at both the radial and angular directions are constant.

2.2 Fourier Descriptors

The Fourier descriptors of a K-point boundary *S(k)*, which can be represented as complex coordinates, centroid distance, curvature signature, and cumulative angular function, are computed by:

$$a(u) = \frac{1}{K}\sum_{k=0}^{K-1} s(k) e^{-\frac{j2\pi uk}{K}} \quad \text{for } u = 0,1,2,\ldots,\text{K-1}. \quad (2)$$

The transformed boundary in the Fourier frequency domain captures general shape information in the lower-frequency coefficients and the shape details in the higher-frequency coefficients.

2.3 Wavelet Transform Modulus Maxima

Wavelet transforms can decompose images into elementary building blocks localized both in space and frequency. This decomposition provides a natural approach for the multi-level image contour analysis. Hwang and Mallat [6] prove a local maximum of the wavelet transform (i.e., a strict local maximum of the modulus on either its right or left side) at the finer scales corresponds to a singularity. This singularity is a measure of the local regularity of the intensities and precisely identifies discontinuities in the intensity in images. Therefore, wavelet transforms are particularly useful in recognition of edges, boundaries and important features. The WTMM can further be utilized for contour description in the curvature analysis. The WTMM-based descriptors, unlike Fourier transformation-based global contour descriptors, provide precise local shape information.

3. Shape Indexing and Retrieval

This section details the step-by-step procedure for shape indexing and retrieval. We focus on a boundary-based representation of shape rather than a region-based one due to lateral inhibition, which behaves like the high-pass Laplacian operator [7]. Every image in the database contains one non-occluded object boundary and is indexed. The signatures of each image are stored in two indexing files, which respectively contain global and local indexing information and are further used for retrieval.

3.1 Shape Indexing: Global Signature Construction

Step 1: Normalize each candidate image.
A normalization step is necessary for removing noise and tiny details along the boundary. This step ensures that the shape boundary of each image has the same number of data points, which is required in the following processing steps. The equal-point sampling technique is chosen to select the NSPs (Normalized Sampling Points). All the consecutive NSPs are spaced at equal number of points along the boundary, which is determined by *L/N*, where *L* is the total number of points along the boundary, and *N* is the number of NSPs. We choose *N* to be 512 for the computational convenience.

Step 2: Apply LPM to each NSP.
LPM is applied on each normalized image using a resolution of *N*-by-*N*. Only the NSPs are involved in the mapping using (1). The origin of the log-polar coordinates is the centroid of the normalized shape.

Step 3: Align each LPMed image.
The alignment is achieved by circularly shifting all the mapped NSPs to the left direction. The shifting offset is determined by the column number of the mapped NSP, which has the largest distance (i.e., *Max_Dist*) to the origin of the log-polar coordinates.

Step 4: Apply ILPM to restore the aligned image to the original Cartesian coordinates.
The largest *Max_Dist* among all the images in the database is utilized by ILPM as the resolution (sampling rate) to restore each aligned image. This restoration eliminates the scaling effect so all the images in the database are in the same scale. Fig. 1 illustrates two images (image (b) is a transformed version of image (a)) and their preprocessed images after applying the above

four steps. It clearly shows that the four preprocessing steps ensure the RST invariant property.

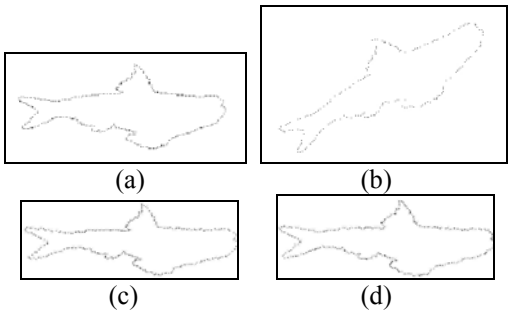


Fig. 1: Final results of the preprocessing steps
(a) Original image
(b) Identical image (enlarge to 1.2 and rotate 30°)
(c) Preprocessed (a)
(d) Preprocessed (b)

Step 5: Construct the global signature by the FDs.
Fourier transformation is applied on the complex coordinates of each restored image. The DC component of the FDs is discarded since it is related to the position of the shape. The invariant global signature is to index the shape and is constructed as:

$$f = \left\{ \frac{|FD_2|}{|FD_2|}, \frac{|FD_3|}{|FD_2|}, ..., \frac{|FD_N|}{|FD_2|} \right\} \quad (3)$$

3.2 Shape Indexing: Local Signature Construction

Step 1: Preprocess each candidate image.
The longest line connecting any two points on the boundary of every image is determined as the reference line. Each image in the database is registered by rotating an angle between its reference and the horizontal lines.

Step 2: Determine the important WTMMs.
A one-level wavelet decomposition is applied to each preprocessed image. The WTMMs in the diagonal subband are located using a 3-by-3 neighborhood. The important WTMMs are determined by setting up a threshold T equal to 30% of the largest WTMM. Fig. 2 demonstrates an image and its WTMM image. It is shown in Fig. 2(b) that the WTMM image captures the major contour information of the original image.

Step 3: Construct the local signature.
The invariant local signature is to index the shape and is constructed as a (Distance, Angle) pair from each important WTMM point to the centroid of the shape starting at the point with the smallest angle to the horizontal direction. The distance is also normalized to a range of 0 and 1 to reduce the scaling effect.

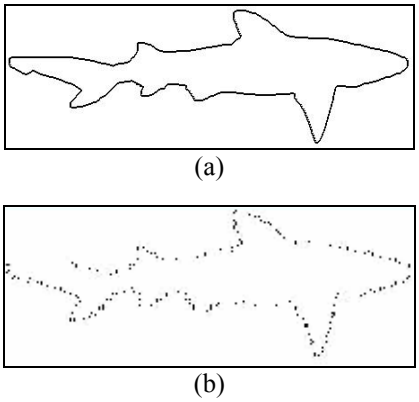


Fig. 2: Illustration of the WTMM image
(a) Original image
(b) Its WTMM image

3.3 Shape Retrieval: Fuzzy Measurement

The global and local signatures of the query image are computed using the same indexing procedure. These two signatures are compared with the corresponding signatures of each candidate image in the database for possible matches.

The Euclidean distance calculates the similarity between the global signatures of the query and candidate images. The smaller the Euclidean distance between two images, the more similar they are.

The similarity score between the local signatures of the query and candidate images is computed:

$$S = \frac{2 \times (K - \xi)}{N_1 + N_2} \times 100 \quad (4)$$

where:

- K is the number of matched maxima, which are determined by the corresponding distance and angle differences that are respectively less than predefined thresholds T_d and T_a;
- N_1 and N_2 are respectively the number of maxima of the query and candidate images;
- ξ measures how good the match is between the query and candidate images and is calculated:

$$\xi = \sum_{i=1}^{K} \left(\frac{\delta D_i}{mean(D_{1i}, D_{2i})} + \frac{\delta A_i}{mean(A_{1i}, A_{2i})} \right) \quad (5)$$

where D_{1i} and D_{2i} are the distances of each matched maxima to the centroid of the query and candidate images; A_{1i} and A_{2i} are the angles of each matched maxima to the centroid of the query and candidate images; δD_{ii} and δA_i are the differences between the query and candidate images in terms of each matched maxima pair (Distance, Angle).

The higher the S value between two images, the more similar they are.

A Gaussian-based fuzzy method is then used to calculate an overall similarity score by combining the global and local similarity measures. The Gaussian function

$$A(x) = e^{-k(x-m)^2} \quad (6)$$

is used to compute the membership for the global and local similarity, where different k's and m's are chosen to evenly distribute each membership between 0 and 1. The average of these two memberships is the final similarity membership between the query and candidate images. λ cut is finally used to defuzzify the final similarity membership and retrieve the candidate images that are most similar to the query image.

4. Experimental Results

The proposed shape retrieval approach is tested on several databases to illustrate its effectiveness.

4.1 Experiments on the RST-Invariance

The first database consists of 180 images and is formed by performing different RST transformations on several distinct objects. Fig. 3 shows the retrieval results from the first database by using a hand as a query image. It clearly demonstrates the proposed approach can successfully retrieve the RST transformed object.

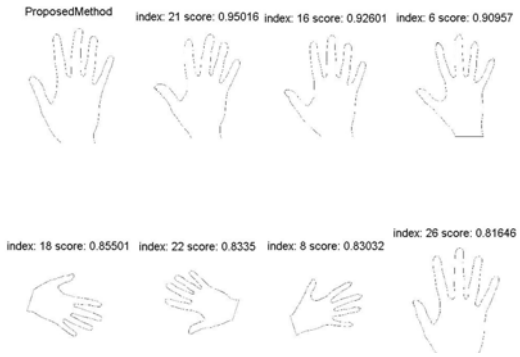
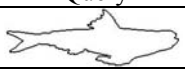
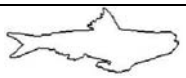
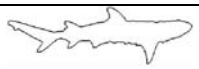
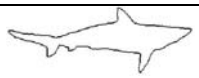
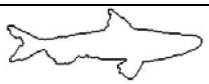
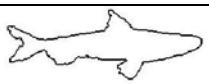


Fig. 3: Top 7 retrieval results from the first database by using a hand as a query

The second database contains 150 images and is formed by performing three RST transformations on each of the 50 randomly chosen images from the SQUID (Shape Query Using Image Databases) [8]. Table 1 shows the retrieval results from the second database by respectively using two images in the second database as a query. The retrieval results are listed starting from the most similar candidate image. The similarity scores are listed below each retrieved image. It is clearly observed that all the three RST transformed images of the query image are returned as the top three retrieval results even though other candidate images are visually similar to the query image as well. All returned top 10 retrieval images resemble the corresponding query image.

Table 1: Top 10 retrieval results from the second database by using two queries

Query 1	Query 2
	
Retrieval Results	Retrieval Results
 Score: 0.9863	 Score: 0.9551
 Score: 0.9435	 Score: 0.9417
 Score: 0.9397	 Score: 0.9337
 Score: 0.9253	 Score: 0.8653
 Score: 0.9134	 Score: 0.8457
 Score: 0.9121	 Score: 0.8208
 Score: 0.9100	 Score: 0.8143
 Score: 0.8999	 Score: 0.8051
 Score: 0.8915	 Score: 0.802
 Score: 0.8900	 Score: 0.7911

4.2 Retrieval Performance Comparisons

Table 2 demonstrates the retrieval results from the third database, which consists of 200 randomly selected images from the SQUID, by applying three different techniques (WTMM, LPM+FDs, and our approach) using the same query image 1 in Table 1. The results from Table 2 illustrate that our proposed method provides better retrieval results than the other two techniques.

Table 2: Retrieval results of three methods

WTMM	LPM+FDs	Proposed Method
 Score:66.9777	 Score: 0.1083	 Score: 0.9036
 Score: 64.0393	 Score: 0.1124	 Score: 0.8852
 Score: 62.9709	 Score: 0.1235	 Score: 0.8651
 Score: 54.6251	 Score: 0.1264	 Score: 0.8252
 Score: 53.4092	 Score: 0.1371	 Score: 0.8238

The fourth database contains 99 shapes and is downloaded from Sebastian's research website. It is used to compare our proposed approach with the method proposed in [9]. This database includes nine categories with 11 shapes in each category. Table 3 shows the number of times for each category that the *n* nearest matches are in the appropriate category. In this experiment, each shape is used as a model to which all the other shapes are compared. As a result, 9801 shape comparisons are made for each repetition of the experiment. The retrieval inaccuracy is mainly caused by the missing or occluded parts of some shapes such as planes, rabbits, and hands. Even though our approach performs a little bit worse than the method proposed in [9] in the second, third, fourth, and fifth nearest matches, the overall recognition performance is 76.4%, which is higher than the 72.7% obtained by [9].

5. Conclusions

We propose a new approach to shape-based image retrieval. It captures both global and local RST invariant shape information by integrating LPM with FDs, and applying the WTMM-based technique on the registered images. A Gaussian-based fuzzy method is further used to calculate the overall similarity using the membership of the global and local similarity measures. This technique is robust to RST changes after several preprocessing. It is also efficient to retrieve images within a large database due to the easy construction of both global and local shape signatures.

Acknowledgements

The authors would like to thank Professors F. Mokhtarian, J. Kittler, and T. Sebastian for providing part of the test images.

References:

[1] D. S. Zhang & G. Lu, A comparative study on shape retrieval using Fourier descriptors with different shape signatures, *Proc. of Int Conf on Intelligent Multimedia and Distance Education*, Fargo, ND, 2001, 1-9.

[2] S. Belongie, J. Malic, & J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans on PAMI, 24(24)*, 2002, 509-522.

[3] F. A. Cheikh, A. Quddus, & M. Gabbouj, Multi-level shape recognition based on wavelet-transform modulus maxima, *4th IEEE Southwest Symposium on Image Analysis and Interpretation*, Austin, TX, 2000, 8-12.

[4] Y. Chen & J. Z. Wang, A region-based fuzzy feature matching approach to content-based image retrieval, *IEEE Trans on PAMI, 24(9)*, 2002, 1252-1267.

[5] D. Zheng , J. Zhao, & A. E. Saddik, RST invariant digital image watermarking based on log-polar mapping and phase correlation, *IEEE Trans on Circuits and Systems for Video Technology,13(8)*, 2003, 753-765.

[6] S. Mallat, & W. L. Hwang, Singularity detection and processing with wavelets, *IEEE Trans on Information Theory*, *38(2)*, 1992, 617-643.

[7] J. C. Russ, *The Image Processing Handbook*, 3rd edition, CRC, Springer and IEEE Press Inc, 1995.

[8] http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html (Fish contour images)

[9] T. Bernier & J. A. Landry, A new method for representing and matching shapes of natural objects, *Pattern Recognition, 36*, 2003,1711-1723.

Table 3: Resulting nearest matches from the fourth database by category

Category	*n*th nearest match										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Fish	11	11	11	11	10	10	10	10	9	7	1
Rabbits	11	10	10	10	10	10	9	7	8	10	3
Planes	11	10	9	9	8	7	7	2	6	2	2
Greebles	11	10	8	10	6	6	5	4	4	5	3
Wrenches	11	11	11	11	11	11	11	11	11	9	7
Hands	11	11	11	11	9	7	6	5	4	6	2
Humans	11	11	11	11	11	11	10	10	7	10	7
Quadrupeds	11	10	8	7	8	7	4	5	3	2	2
Rays	11	11	11	10	11	11	11	11	11	11	4
Ours: %	100.0	96.0	90.9	90.9	84.8	80.8	73.7	65.7	63.6	62.6	31.3
[9]: %	100.0	98	93.9	92.9	85.9	76.8	70.7	65.7	53.5	37.4	25.3